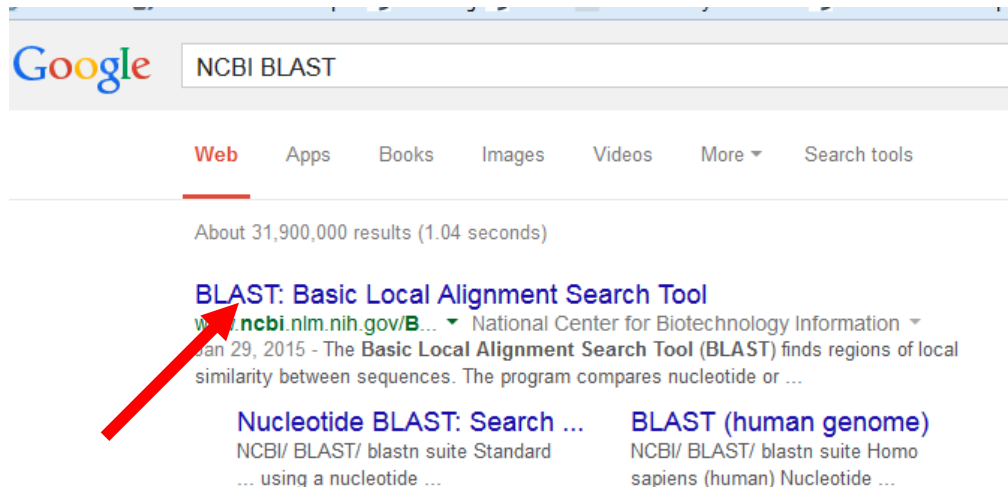On page 135 of "The Lost World", Michael Crichton's fantasy about cloning dinosaurs is the following sequence:

```
GAATTCCGGAAGCGAGCAAGAGATAAGTCCTGGCATCAGATACAGTTGGAGATAAGGACG
GACGTGTGGCAGCTCCCGCAGAGGATTCACTGGAAGTGCATTACCTATCCCATGGGAGCC
ATGGAGTTCGTGGCGCTGGGGGGGCCGGATGCGGGCTCCCCCACTCCGTTCCCTGATGAA
GCCGGAGCCTTCCTGGGGCTGGGGGGGGGCGAGAGGACGGAGGCGGGGGGGGCTGCTGGCC
TCCTACCCCCCCTCAGGCCGCGTGTCCCTGGTGCCGTGGGCAGACACGGGTACTTTGGGG
ACCCCCCAGTGGGTGCCGCCCGCCACCCAAATGGAGCCCCCCCACTACCTGGAGCTGCTG
CAACCCCCCCGGGGCAGCCCCCCCCATCCCTCCTCCGGGCCCCTACTGCCACTCAGCAGC
GGGCCCCCACCCTGCGAGGCCCGTGAGTGCGTCATGGCCAGGAAGAACTGCGGAGCGACG
GCAACGCCGCTGTGGCGCCGGGACGGCACCGGGCATTACCTGTGCAACTGGGCCTCAGCC
TGCGGGCTCTACCACCGCCTCAACGGCCAGAACCGCCCGCTCATCCGCCCCAAAAAGCGC
CTGCTGGTGAGTAAGCGCGCAGGCACAGTGTGCAGCCACGAGCGTGAAAACTGCCAGACA
TCCACCACCACTCTGTGGCGTCGCAGCCCCATGGGGGACCCCGTCTGCAACAACATTCAC
GCCTGCGGCCTCTACTACAAACTGCACCAAGTGAACCGCCCCCCTCACGATGCGCAAAGAC
GGAATCCAAACCCGAAACCGCAAAGTTTCCTCCAAGGGTAAAAAGCGGCGCCCCCCGGGG
GGGGGAAACCCCTCCGCCACCGCGGGAGGGGGCGCTCCTATGGGGGGAGGGGGGGACCCC
TCTATGCCCCCCCCGCCGCCCCCCCCGGCCGCCGCCCCCCCTCAAAGCGACGCTCTGTAC
GCTCTCGGCCCCGTGGTCCTTTCGGGCCATTTTCTGCCCTTTGGAAACTCCGGAGGGTTT
TTTGGGGGGGGGGCGGGGGGTTACACGGCCCCCCCGGGGCTGAGCCCGCAGATTTAAATA
ATAACTCTGACGTGGGCAAGTGGGCCTTGCTGAGAAGACAGTGTAACATAATAATTTGCA
CCTCGGCAATTGCAGAGGGTCGATCTCCACTTTGGACACAACAGGGCTACTCGGTAGGAC
CAGATAAGCACTTTGCTCCCTGGACTGAAAAGAAAGGATTTATCTGTTTGCTTCTTGCT
GACAAATCCCTGTGAAAGGTAAAAGTCGGACACAGCAATCGATTATTTCTCGCCTGTGTG
AAATTACTGTGAATATTGTAAATATATATATATATATATATATATCTGTATAGAACAGCC
TCGGAGGCGGCATGGACCCAGCGTAGATCATGCTGGATTTGTACTGCCGGAATTC
```

We will use BLAST to check if the sequence really belongs to a dinosaur. Here are the steps:

1. Copy the sequence.
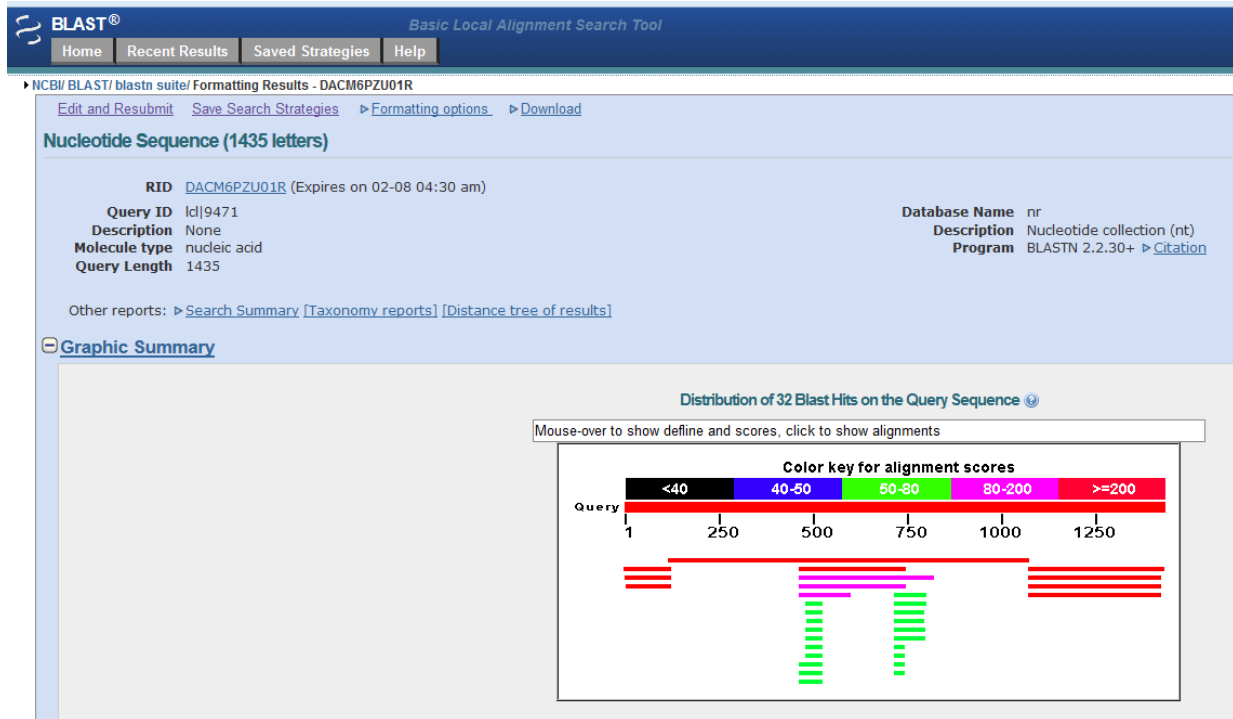2. Search with Google for: *NCBI BLAST*

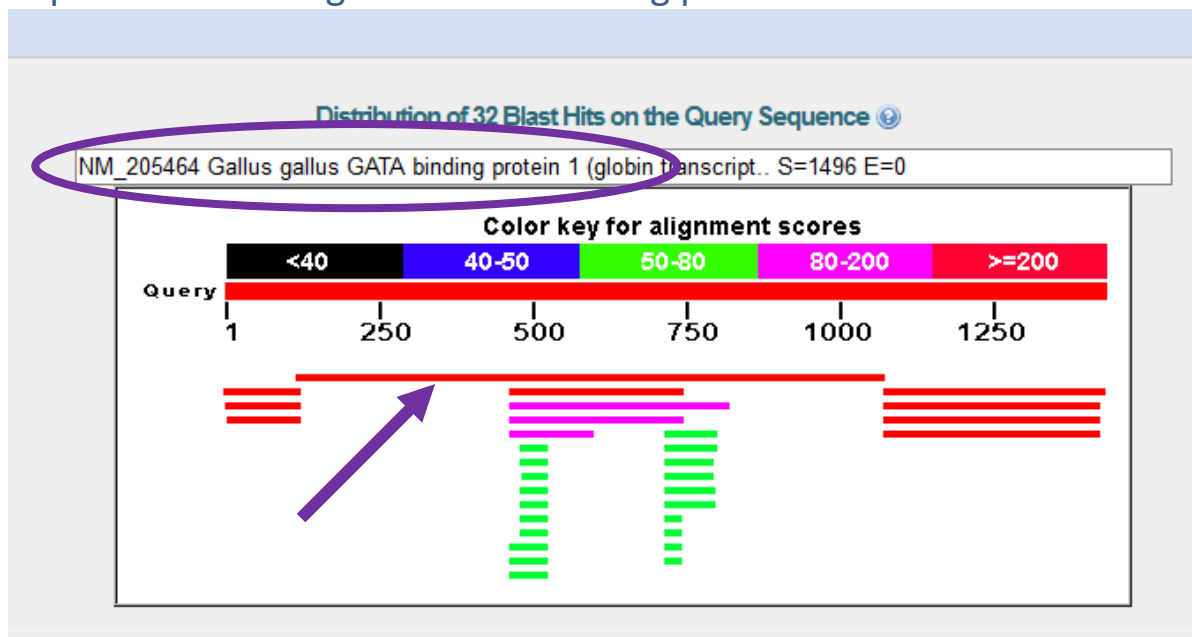3. Select the BLAST link that takes you to the BLAST home page

4. On the BLAST home page select the link to the "nucleotide blast".

NCBI/ BLAST/ blastn suite   **Standard Nucleotide BLAST**

blastn   blastp   blastx   tblastn   tblastx

Enter Query Sequence   BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter accession number(s), gi(s), or FASTA sequence(s) ⓘ   Clear   Query subrange ⓘ

```
GAATTCCGGAAGCGAGCAAGAGATAAGTCCTGGCATCAGATACAGTTGGAGATAAGGACGGACGTGTGGC
AGCTCCCGCAGAGGATTCACTGGAAGTGCATTACCTATCCCATGGGAGCCATGGAGTTCGTGGCGCTGGG
GGGGCCGGATGCGGGCTCCCCCACTCCGTTCCCTGATGAAGCCGGAGCCTTCCTGGGGCTGGGGGGGGGC
GAGAGGACGGAGGCGGGGGGGCTGCTGGCCTCCTACCCCCCCTCAGGCCGCGTGTCCCTGGTGCCGTGGG
CAGACACGGGTACTTTGGGGACCCCCCAGTGGGTGCCGCCCGCCACCCAAATGGAGCCCCCCCACTACCT
GGAGCTGCTGCAACCCCCCCGGGGCAGCCCCCCCCATCCCTCCTCCGGGCCCCTACTGCCACTCAGCAGC
GGGCCCCCACCCTGCGAGGCCCGTGAGTGCGTCATGGCCAGGAAGAACTGCGGAGCGACGGCAACGCCGC
TGTGGCGCCGGGACGGCACCGGGCATTACCTGTGCAACTGGGCCTCAGCCTGCGGGCTCTACCACCGCCT
CAACGGCCAGAACCGCCCGCTCATCCGCCCCAAAAAGCGCCTGCTGGTGAGTAAGCGCGCAGGCACAGTG
TGCAGCCACGAGCGTGAAAACTGCCAGACATCCACCACCACTCTGTGGCGTCGCAGCCCCATGGGGGACC
CCGTCTGCAACAACATTCACGCCTGCGGCCTCTACTACAAACTGCACCAAGTGAACCGCCCCCTCACGAT
GCGCAAAGACGGAATCCAAACCCGAAACCGCAAAGTTTCCTCCAAGGGTAAAAAGCGGCGCCCCCCGGGG
GGGGGAAACCCCTCCGCCACCGCGGGAGGGGGCGCTCCTATGGGGGGAGGGGGGGACCCCTCTATGCCCC
CCCCGGTTTCCCCCCCCGGCCGCCGCCCCCCCTCAAAGCGACGCTCTGTACGCTCTCGGCCCCGTGGTCCT
TTCGTTTCATTTCTGCCCTTTGGAAACTCCGGAGGGTTTTTTGGGGGGGGGCGGGGGGTTACACGGGCC
CCCCGGGGCTGAGCCCGCAGATTTAAATAATAACTCTGACGTGGGCAAGTGGGCCTTGCTGAGAAGACA
AGTAACATAATAATTTGCACCTCGGCAATTGCAGAGGGTCGATCTCCACTTTGGACACAACAGGGCTAC
TCGGTAGGACCAGATAAGCACTTTGCTCCCTGGACTGAAAAAGAAAGGATTTATCTGTTTGCTTCTTGCT
GACAAATCCCTGTGAAAGGTAAAAGTCGGACACAGCAATCGATTATTTCTCGCCTGTGTGAAATTACTGT
GAATATTGTAAATATATATATATATATATATATATCTGTATAGAACAGCCTCGGAGGCGGCATGGACCCA
GCGTAGATCATGCTGGATTTGTACTGCCGGAATTC
```

From [   ]
To [   ]

Or, upload file   [Browse…] No file selected. ⓘ

Job Title   [Nucleotide Sequence (1435 letters)   ]
             Enter a descriptive title for your BLAST search ⓘ

☐ Align two or more sequences ⓘ

5. On the standard Nucleotide BLAST page enter your query sequence.


```
CCCCCGGGGCTGAGCCCGCAGATTTAAATAATAACTCTGACGTGGGCAAGTGGGCCTTGCTGAGAAGACA
GTGTAACATAATAATTTGCACCTCGGCAATTGCAGAGGGTCGATCTCCACTTTGGACACAACAGGGCTAC
TCGGTAGGACCAGATAAGCACTTTGCTCCCTGGACTGAAAAAGAAAGGATTTATCTGTTTGCTTCTTGCT
GACAAATCCCTGTGAAAGGTAAAAGTCGGACACAGCAATCGATTATTTCTCGCCTGTGTGAAATTACTGT
GAATATTGTAAATATATATATATATATATATATATCTGTATAGAACAGCCTCGGAGGCGGCATGGACCCA
GCGTAGATCATGCTGGATTTGTACTGCCGGAATTC
```

Or, upload file   [Browse…] No file selected. ⓘ

Job Title   [Nucleotide Sequence (1435 letters)   ]
             Enter a descriptive title for your BLAST search ⓘ

☐ Align two or more sequences ⓘ

Choose Search Set

Database   ○ Human genomic + transcript ○ Mouse genomic + transcript ● Others (nr etc.):
              [Nucleotide collection (nr/nt)   ▼] ⓘ

Organism   [Enter organism name or id--completions will be suggested   ] ☐ Exclude [+]
Optional   Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ⓘ

Exclude   ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences
Optional

Limit to   ☐ Sequences from type material
Optional

Entrez Query [   ] [You Tube] Create custom database
Optional   Enter an Entrez query to limit search ⓘ

Program Selection

Optimize for ● Highly similar sequences (megablast)
              ○ More dissimilar sequences (discontiguous megablast)
              ○ Somewhat similar sequences (blastn)
              Choose a BLAST algorithm ⓘ

[**BLAST**]   Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
                ☐ Show results in a new window

6. There are several settings and options, but you do not have to change any. Simply click on the BLAST button.
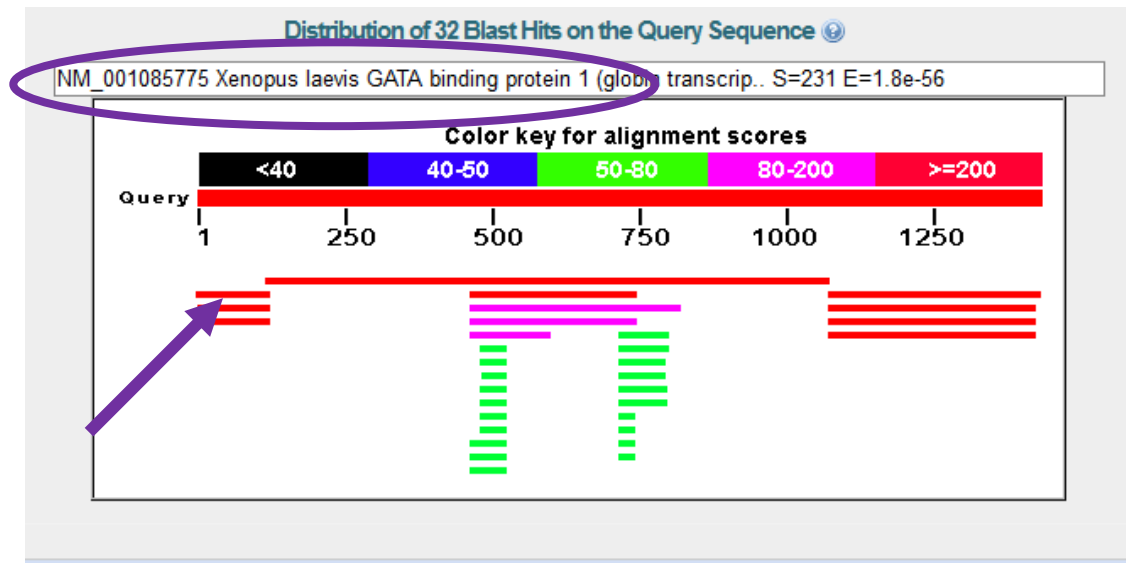
After your BLAST run completes, you should see results like this:

If you mouse over the graphics of the top alignment, it says that the sequence is Gallus gallus GATA binding protein 1.

Similarly, the second alignment is Xenopus laevis GATA binding protein 2 (gata2):

Scroll down the search results page to see the alignment of your query sequence to the Gallus gallus sequence. Link to the subject sequence in GenBank (which is a part of the Nucleotide database). The link opens in a separate tab and you can verify that Gallus gallus is chicken.

Go back to the search results page and check the alignment carefully. The sequences appear identical, except that our query sequence has some extra bases that are not there in the chicken sequence (dashes in the alignment:

Gallus gallus GATA binding protein 1 (globin transcription factor 1) (GATA1), mRNA
Sequence ID: ref|NM_205464.1| Length: 1068 Number of Matches: 1
▷ See 1 more title(s)

Range 1: 39 to 953 GenBank Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1496 bits(810)	0.0	915/957(96%)	42/957(4%)	Plus/Plus

```
Query  119   CCATGGAGTTCGTGGCGCTgggggggCCGGATGCGGGCTCCCCCACTCCGTTCCCTGATG   178
             |||||||||||||||||||      |||||||||||||||||||||||||||||||||||
Sbjct  39    CCATGGAGTTCGTGGCGCTGGGGGGGGCCGGATGCGGGCTCCCCCACTCCGTTCCCTGATG   98

Query  179   AAGCCGGAGCCTTCCTGGGGCTggggggggggCGAGAGGACGGAGGCgggggggCTGCTGG   238
             |||||||||||||||||||||          |||||||||||||||       |||||||
Sbjct  99    AAGCCGGAGCCTTCCTGGGGCTGGGGGGGGGCGAGAGGACGGAGGCGGGGGGGCTGCTGG   158

Query  239   CCTCCTAccccccccTCAGGCCGCGTGTCCCTGGTGCCGTGGGCAGACACGGGTACTTTGG   298
             |||||||       |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  159   CCTCCTACCCCCCCCTCAGGCCGCGTGTCCCTGGTGCCGTGGGCAGACACGGGTACTTTGG   218

Query  299   GGACCCCCCAGTGGGTGCCGCCCGCCACCCAAATGGAGccccccccACTACCTGGAGCTGC   358
             |||||||||||||||||||||||||||||||||||||        ||||||||||||||||
Sbjct  219   GGACCCCCCAGTGGGTGCCGCCCGCCACCCAAATGGAGCCCCCCCACTACCTGGAGCTGC   278

Query  359   TGCAAccccccggggcagcccccccccATCCCTCCTCCGGGCCCCTACTGCCACTCAGCA   418
             |||||           |         |||||||||||||||||||||||||||||||||||
Sbjct  279   TGCAACCCCCCCGGGGCAGCCCCCCCCCATCCCTCCTCCGGGCCCCTACTGCCACTCAGCA   338

Query  419   GCGGGCCCCCACCCTGCGAGGCCCGTGAGTGCGTCATGGCCAGGAAGAACTGCGGGAGCGA   478
             |||||||||||||||||||||||||||||||||||            ||||||||||||||
Sbjct  339   GCGGGCCCCCACCCTGCGAGGCCCGTGAGTGCGTC------------AACTGCGGGAGCGA   386

Query  479   CGGCAACGCCGCTGTGGCGCCGGGACGGCACCGGGCATTACCTGTGCAACTGGGCCTCAG   538
             |||||||||||||||||||||||||||||||||||||||||||||||||         |
Sbjct  387   CGGCAACGCCGCTGTGGCGCCGGGACGGCACCGGGCATTACCTGTGCAAC---------G   437

Query  539   CCTGCGGGCTCTACCACCGCCTCAACGGCCAGAACCGCCCGCTCATCCGCCCCAAAAAGC   598
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  438   CCTGCGGGCTCTACCACCGCCTCAACGGCCAGAACCGCCCGCTCATCCGCCCCAAAAAGC   497

Query  599   GCCTGCTGGTGAGTAAGCGCGCAGGCACAGTGTGCAGCCACGAGCGTGAAAACTGCCAGA   658
             ||||||||||||||||||||||||||||||||||||||            ||||||||||
Sbjct  498   GCCTGCTGGTGAGTAAGCGCGCAGGCACAGTGTGCAGC------------AACTGCCAGA   545

Query  659   CATCCACCACCACTCTGTGGCGTCGCAGCCCCATGGGGGACCCCGTCTGCAACAACATTC   718
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  546   CATCCACCACCACTCTGTGGCGTCGCAGCCCCATGGGGGACCCCGTCTGCA---------   596

Query  719   ACGCCTGCGGCCTCTACTACAAACTGCACCAAGTGAACCGCCCCCTCACGATGCGCAAAG   778
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  597   ACGCCTGCGGCCTCTACTACAAACTGCACCAAGTGAACCGCCCCCTCACGATGCGCAAAG   656

Query  779   ACGGAATCCAAACCCGAAACCGCAAAGTTTCCTCCAAGGGTAAAAAGCGGCGCCCCCCgg   838
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  657   ACGGAATCCAAACCCGAAACCGCAAAGTTTCCTCCAAGGGTAAAAAGCGGCGCCCCCCGG   716
```

Go back to very top of the BLAST page where you can change the "Formatting options" to show the "CDS feature" and "Reformat". This will show how the DNA sequence translates in amino acid sequence (there will be different letters each representing a different amino acid sequence):

Range 1: 39 to 953 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
1496 bits(810)	0.0	915/957(96%)	42/957(4%)	Plus/Plus

```
CDS: Putative 1    1       M  E  F  V  A  L  G  G  P  D  A  G  S  P  T  P  F  P  D
Query            119       CCATGGAGTTCGTGGCGCTgggggggCCGGATGCGGGCTCCCCCACTCCGTTCCCTGATG   178
                           |||||||||||||||||||      |||||||||||||||||||||||||||||||||||
Sbjct             39       CCATGGAGTTCGTGGCGCTGGGGGGGGCCGGATGCGGGCTCCCCCACTCCGTTCCCTGATG    98
CDS:erythroid transc 1     M  E  F  V  A  L  G  G  P  D  A  G  S  P  T  P  F  P  D

CDS: Putative 1   20       E  A  G  A  F  L  G  L  G  G  G  E  R  T  E  A  G  G  L  L
Query            179       AAGCCGGAGCCTTCCTGGGGCTggggggggggCGAGAGGACGGAGGCgggggggCTGCTGG   238
                           |||||||||||||||||||||             ||||||||||||||         |||||||
Sbjct             99       AAGCCGGAGCCTTCCTGGGGCTGGGGGGGGGCGAGAGGACGGAGGCGGGGGGGCTGCTGG   158
CDS:erythroid transc 20    E  A  G  A  F  L  G  L  G  G  G  E  R  T  E  A  G  G  L  L

CDS: Putative 1   40       A  S  Y  P  P  S  G  R  V  S  L  V  P  W  A  D  T  G  T  L
Query            239       CCTCCTAcccccccTCAGGCCGCGTGTCCCTGGTGCCGTGGGCAGACACGGGTACTTTGG   298
                           |||||||       |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct            159       CCTCCTACCCCCCCTCAGGCCGCGTGTCCCTGGTGCCGTGGGCAGACACGGGTACTTTGG   218
CDS:erythroid transc 40    A  S  Y  P  P  S  G  R  V  S  L  V  P  W  A  D  T  G  T  L

CDS: Putative 1   60       G  T  P  Q  W  V  P  P  A  T  Q  M  E  P  P  H  Y  L  E  L
Query            299       GGACCCCCCAGTGGGTGCCGCCCGCCACCCAAATGGAGcccccccACTACCTGGAGCTGC   358
                           ||||||||||||||||||||||||||||||||||||||       |||||||||||||||
Sbjct            219       GGACCCCCCAGTGGGTGCCGCCCGCCACCCAAATGGAGCCCCCCCACTACCTGGAGCTGC   278
CDS:erythroid transc 60    G  T  P  Q  W  V  P  P  A  T  Q  M  E  P  P  H  Y  L  E  L

CDS: Putative 1   80       L  Q  P  P  R  G  S  P  P  H  P  S  S  G  P  L  L  P  L  S
Query            359       TGCAAcccccccgggggcagcccccccccATCCCTCCTCCGGGCCCCTACTGCCACTCAGCA   418
                           |||||       ||||||||      |||||||||||||||||||||||||||||||||||
Sbjct            279       TGCAACCCCCCCCGGGGCAGCCCCCCCCCATCCCTCCTCCGGGCCCCTACTGCCACTCAGCA   338
CDS:erythroid transc 80    L  Q  P  P  R  G  S  P  P  H  P  S  S  G  P  L  L  P  L  S

CDS: Putative 1  100       S  G  P  P  P  C  E  A  R  E  C  V  [M  A  R  K  N] C  G  A
Query            419       GCGGGCCCCCACCCTGCGAGGCCCGTGAGTGCGTCATGGCCAGGAAGAACTGCGGAGCGA   478
                           |||||||||||||||||||||||||||||||||||            |||||||||||||
Sbjct            339       GCGGGCCCCCACCCTGCGAGGCCCGTGAGTGCGTC------------AACTGCGGAGCGA   386
CDS:erythroid transc 100   S  G  P  P  P  C  E  A  R  E  C  V              N  C  G  A

CDS: Putative 1  120       T  A  T  P  L  W  R  R  D  G  T  G  H  Y  L  C  N  [W  A  S]
Query            479       CGGCAACGCCGCTGTGGCGCCGGGACGGCACCGGGCATTACCTGTGCAACTGGGCCTCAG   538
                           |||||||||||||||||||||||||||||||||||||||||||||||||||          |
Sbjct            387       CGGCAACGCCGCTGTGGCGCCGGGACGGCACCGGGCATTACCTGTGCAAC---------G   437
CDS:erythroid transc 116   T  A  T  P  L  W  R  R  D  G  T  G  H  Y  L  C  N

CDS: Putative 1  140       A  C  G  L  Y  H  R  L  N  G  Q  N  R  P  L  I  R  P  K  K
Query            539       CCTGCGGGCTCTACCACCGCCTCAACGGCCAGAACCGCCCGCTCATCCGCCCCAAAAAGC   598
                           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct            438       CCTGCGGGCTCTACCACCGCCTCAACGGCCAGAACCGCCCGCTCATCCGCCCCAAAAAGC   497
CDS:erythroid transc 133   A  C  G  L  Y  H  R  L  N  G  Q  N  R  P  L  I  R  P  K  K

CDS: Putative 1  160       R  L  L  V  S  K  R  A  G  T  V  C  S  [H  E  R  E] N  C  Q
Query            599       GCCTGCTGGTGAGTAAGCGCGCAGGCACAGTGTGCAGCCACGAGCGTGAAAACTGCCAGA   658
                           ||||||||||||||||||||||||||||||||||||||            ||||||||||
Sbjct            498       GCCTGCTGGTGAGTAAGCGCGCAGGCACAGTGTGCAGC------------AACTGCCAGA   545
CDS:erythroid transc 153   R  L  L  V  S  K  R  A  G  T  V  C  S              N  C  Q

CDS: Putative 1  180       T  S  T  T  T  L  W  R  R  S  P  M  G  D  P  V  C  N  [N  I]
Query            659       CATCCACCACCACTCTGTGGCGTCGCAGCCCCATGGGGGACCCCGTCTGCAACAACATTC   718
                           |||||||||||||||||||||||||||||||||||||||||||||||||||          
Sbjct            546       CATCCACCACCACTCTGTGGCGTCGCAGCCCCATGGGGGACCCCGTCTGCA---------   596
CDS:erythroid transc 169   T  S  T  T  T  L  W  R  R  S  P  M  G  D  P  V  C

CDS: Putative 1  200       [H] A  C  G  L  Y  Y  K  L  H  Q  V  N  R  P  L  T  M  R  K
Query            719       ACGCCTGCGGCCTCTACTACAAACTGCACCAAGTGAACCGCCCCCCTCACGATGCGCAAAG   778
                           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct            597       ACGCCTGCGGCCTCTACTACAAACTGCACCAAGTGAACCGCCCCCCTCACGATGCGCAAAG   656
CDS:erythroid transc 186   N  A  C  G  L  Y  Y  K  L  H  Q  V  N  R  P  L  T  M  R  K
```

The "dinosaur" sequence really is constructed from sequences of chicken and frog genes. It was provided to Michael Crichton by Mark Boguski who used to work at NIH. Mark added (made up) several bases that when translated into the one letter codes of amino acids spell out as: MARK WAS HERE NIH.